

# Query Method of Web Education Resources Based on Semantic Topic Similarity

**Yan Shi**

Zhoukou Vocational and Technical College, Zhoukou, Henan, China

## **Abstract**

This paper researched on the query method of web education resources based on semantic topic similarity. This paper proposes a hierarchical model of web Education. This model is a two-tier filtering model based on mobile agent. The model first classifies and filters according to semantic similarity, and then filters web education resources according to Q-learning filtering algorithm in machine learning. This paper presents a kind of rank algorithm for web educational resources. According to Euclid ambiguity and RSS aggregation technology, this algorithm finds the web education resources that users need. First of all, RSS document aggregation technology is used to quickly gather the music education resources that users need. Secondly, the query is characterized by Euclid fuzziness in fuzzy sets. The fuzziness of the association between content and resources. Finally, we get the personalized resources to satisfy the end users of web education resources.

**Keywords:** Web Education, Resource pre selection, Hierarchical model, Semantic similarity

## **1. Introduction**

With the rapid development of computer and network technology, the information management of web educational resources has gradually changed from traditional paper management to automatic management [1]. The emergence of web-based web educational resources will effectively improve the management level and service ability of web educational resources. Many web educational resources will be integrated into the overall educational cloud platform, and the web educational resources will truly realize informationization, networking and paperless management [2].

The theory and technology of Agent originates from distributed artificial intelligence, which provides a new way of thinking for solving complex, dynamic and distributed intelligent applications. Some individuals in the society can get the solution of the problem after negotiation, and these individuals are Agent. Agent network system is called Multi-Agent System (MAS) [3-5]. Wooldridge and Jennings think that agents can accomplish their goals autonomously, which means that agents have the characteristics of autonomy, intelligence and reaction. Mobile Agent not only has the characteristics of general Agent, but also can move freely in heterogeneous software and hardware network environment. Educational resource system needs to aggregate multiple available resources to meet the dynamic needs of users and the dynamic changes of environment. The distributed decision-making, decentralized control and coordination capabilities provided by mobile Agent technology can meet the needs of dynamically building resources.

## **2. Hierarchical Model for Pre-selection of Web Educational Resources Based on Mobile Agent**

### **2.1 Filtering algorithm based on semantic similarity**

Resource semantic similarity calculation is to calculate the similarity between the internal meanings of two resources. It has been widely used in information integration, information recommendation and filtering, data mining and other fields, and has become a hot spot in information technology research. There are many different methods to calculate similarity, such as cosine formula, Pearson correlation coefficient, conditional probability

and so on. The following definitions are given for the calculation of semantic similarity of web educational resources.

Define 1 set of web educational resources

Set a group of web education resources with different types and versions as a group of resource filtering objects [6].

$$E = \{e_1, e_2, \dots, e_i, \dots, e_n\}$$

Define 2 Web Education Resource Vector

In the Web education resource  $e$ , feature extraction is carried out on the Web education resource to obtain the resource vector  $V_i = \{w_1, w_2, \dots, w_i, \dots, w_n\}$  with the feature set  $F = \{f_1, f_2, \dots, f_i, \dots, f_n\}$  and  $e_i$ . [7-8]

$$w_i = \alpha \times \frac{f(f_i, e_{th})}{\sum_{i=1}^n f(f_i, e)} + \beta \frac{f(f_i, e_{td})}{\sum_{i=1}^n f(f_i, e)} \quad (1)$$

Definition 3 Semantic Similarity of web educational resources

For resources  $e_1$  and  $e_2$  in the Web education resource set, define

$$sim(e_1, e_2) = \frac{\sum_{i=1}^n w_{1i} \times w_{2i}}{\sqrt{\left(\sum_{i=1}^n w_{1i}^2\right) \left(\sum_{i=1}^n w_{2i}^2\right)}} \quad (2)$$

Firstly, a filtering algorithm HA\_SA based on semantic similarity is adopted, which mainly considers two aspects:

① How to calculate the weight of secondary feature vectors in resources according to the feature vectors extracted from resources; ② How to calculate the semantic similarity between two web educational resources to filter the resources. Due to the heterogeneity and magnanimity of web educational resources, the problem becomes more and more complicated [9-10].

The algorithm uses vector space model to filter Web education resources, quantizes resources into a set of feature vectors, counts the word frequency of feature vectors in resource names (titles) and texts respectively, and calculates the weight of this feature vector. Finally, the similarity between two Web education resources is calculated according to cosine formula. The execution flow and pseudo-code of the algorithm are shown in Figure 1. In which  $s$  represents  $\varepsilon$  given lower limit of similarity, and if the similarity of a resource is lower than this value, the resource is filtered out in this resource class.

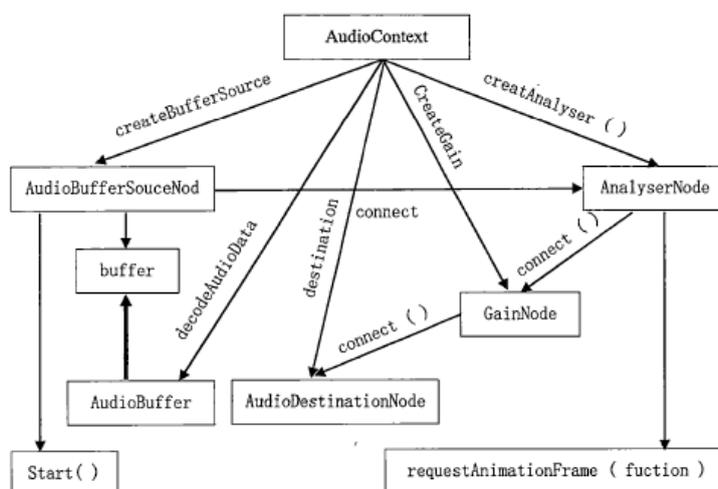


Fig 1: HA\_SA algorithm execution flow

## 2.2 Filtering Algorithm Based on Q Learning

As an unguided machine learning method, reinforcement learning is widely used to model animal and human behavior. In this paper, the users of web educational resources are regarded as the main body of learning, and the environment of web educational resources is regarded as the learning environment, and the web educational resources are filtered through continuous learning feedback. The principle of reinforcement learning is that the Agent performs certain actions on the environment, changes the state of the environment and obtains the return signal given by the environment to strengthen the mapping relationship between a certain state and the optimal action strategy. By repeating this process, the Agent can obtain the ability to give the optimal action strategy in any environmental state.

Property 1 A single Agent MDP contains four elements:  $S, A, P, R$ ,  $S$  is the Agent's state set,  $A$  is the Agent's action set,  $P$  is the state transition function, and  $R$  is the return function.

The learning process of each Agent is described as a quadruple  $(A, R, N, \lambda)$ , where  $a$  represents action,  $r$  represents return value,  $n$  represents satisfaction degree, and threshold value represents the degree of learning process. In the learning process of the user Agent, the return value is continuously returned, and the filtering mechanism of resources is determined according to the size of the return value. In this paper, the traditional Q learning algorithm is used to learn the subject.

The strategy of Q learning algorithm is to estimate the value function (Q value) of  $(s, a)$  by using state-action. In the process of Q learning, Agent learns according to Q function without waiting for the task to be completed.

Update the formula as follows:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha [r_{t+1} + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t)] \quad (3)$$

Where, the parameter  $\alpha$  is the learning step,  $\gamma$  is the discount factor,  $\alpha, \gamma \in (0, 1)$ , and  $r$  is the reward value returned by the user.

The model in this paper is divided into two layers, which are refined from bottom to top step by step, and the first layer is filtered through the calculation of semantic similarity of web educational resources, and several subcategories of web educational resources are obtained. In each subclass, the learning ability of Agent is used to give real-time feedback to users' needs, so as to re-screen the subclass of web educational resources, and finally reduce the scale and complexity of candidate web educational resources, which can better meet users' needs.

Secondly, this paper puts forward a filtering algorithm HA\_Q based on single Agent Q learning, which mainly considers the following three aspects: ① How to adjust the Q value (action-state) of Web education resource user Agent according to learning; ② How to get the return value; ③ How to determine the next action according to the current state. Due to the complexity and mass of users of web educational resources, the problem becomes more and more complicated.

This algorithm uses Q-learning method in the enhanced learning model to filter web education resources, and takes web education resource users as agents, selects appropriate actions to learn continuously according to the Q value (status action) of the initial agent until the learning level of the agent is expected to reach the target. The execution flow and pseudo code of the algorithm are shown in Figure 2. Where  $\gamma$  represents the learning level expected to be reached by the user agent,  $N_i$  is the satisfaction degree. If the satisfaction is lower than the learning level, the state migration is carried out according to Boltzman function, and the resources selected by the user agent in this behavior action are filtered out.

This algorithm uses vector space model to filter web education resources, quantifies resources into a set of feature vectors, calculates the word frequency of feature vectors in resource name (title) and text respectively, and calculates the weight of the feature vector. Finally, the similarity between the two web education resources is calculated according to cosine formula. The execution flow and pseudo code of the algorithm are shown in Figure 2. It indicates the lower limit of the given similarity degree. If the similarity of the resource is lower than this value, the resource will be filtered out in the resource class.

	Conventional office assumptions	New ways of working
Patterns of Work	<ul style="list-style-type: none"> <li>• Routine processes</li> <li>• Individual tasks</li> <li>• Isolated work</li> </ul>	<ul style="list-style-type: none"> <li>• Creative knowledge work</li> <li>• Groups, teams, projects</li> <li>• Interactive work</li> </ul>
Patterns of space occupancy over time	<ul style="list-style-type: none"> <li>• Central office locations</li> <li>• Staff are assumed to occupy individually 'owned' workstations on a full-time basis</li> <li>• Typically over the course of a 9-5 day</li> <li>• The office assumes one desk per person</li> <li>• Provides a hierarchy (planned or enclosed)</li> <li>• Occupied typically at levels at least 30% below full capacity.</li> </ul>	<ul style="list-style-type: none"> <li>• Creative knowledge work</li> <li>• Groups, teams, projects</li> <li>• Interactive work</li> </ul>
Type of space layout, furniture systems and use of space and buildings	<ul style="list-style-type: none"> <li>• Hierarchy of space and furniture related to status.</li> <li>• Individual allocation of space predominates over interactive meeting spaces.</li> </ul>	<ul style="list-style-type: none"> <li>• Distributed set of work locations (which may be nomadic, mobile, office or home)</li> <li>• Linked by a network of communication in which autonomous individuals work in project teams.</li> <li>• Daily timetable is extended and irregular.</li> <li>• Multi-functional work settings are occupied on an as-needed basis.</li> <li>• Daily occupation of space near to capacity.</li> </ul>
Use of information technology	<ul style="list-style-type: none"> <li>• Technology used for routine data processing</li> <li>• Terminals in fixed positions and served by mainframes.</li> </ul>	<ul style="list-style-type: none"> <li>• Focus on mobility of IT equipment</li> <li>• Used in a wide variety of settings.</li> <li>• Technology used to support creative knowledge work (individual and group)</li> <li>• File servers serve a variety of IT tools, including PCs, laptops, shared equip.</li> </ul>

Fig 2: HA\_Q algorithm execution flow

### 3. Prototype system for automatic acquisition of web educational resources

#### 3.1 Experimental environment

- (1) website development: C#
- (2) Development platform: VS2008, SQL Server database

#### 3.2 System architecture

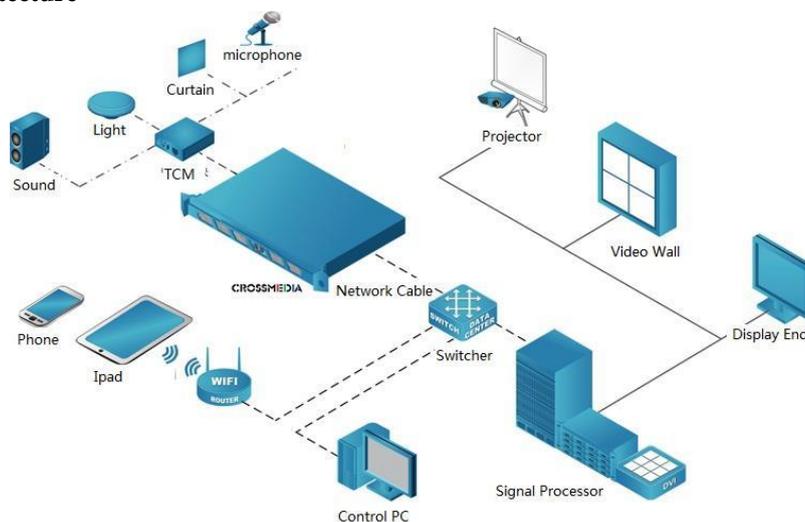


Fig 3: Architecture diagram

#### (1) User registration module

The information registered by users includes user name, password, security information setting, password protection question, password protection question answer, etc. There is also the research on the user's interest in web education resources. In order to facilitate our semantic analysis, we can analyze and recommend the web education resources that users are interested in. User registration will be divided into anonymous login and membership login two large sections.

Anonymous login: when there is anonymous login in the system, we will push resources similar to the search resources. The first step for member login is to register the user, that is, the user information input. It contains user name, age, specialty, interest, basic information of the industry, from which semantic analysis will

automatically push the latest web educational resources of interest to members after login. It also includes a personal resource management section, in order to enable members to manage their own personal resources, and to see the number of times that personal resources are accessed and downloaded in time. The third user login analysis mainly includes the analysis of the number of daily and monthly login and the visit period, so as to provide data proof for the bonus points in the points management. Finally, the user security design includes the password setting, user login, logout and so on, so that users can access the system safely on the premise of confidentiality.

#### (2) Points management module

In order to prevent the free riding phenomenon, we use the points management system to manage the users' behavior of uploading and downloading resources. In order to enable users to provide resources for each other to share, but also for the system to constantly update resources to contribute. Points management is divided into three parts: bonus rules, penalty rules and user permissions. The rules are as follows:

1. Bonus rules: mainly for members, then there are certain incentive policies for members who upload resources and the number of times personal resources are visited and downloaded.
2. Penalty rules: there will be a penalty policy for members who download resources.
3. User rights: when the score reaches a certain amount, it will be upgraded to VIP user, and when the penalty is reduced to a certain amount, it will also be degraded to ordinary user. The difference between VIP users and ordinary users is that they pay different points when using resources.

#### (3) Service Customization

Mainly based on the member login or anonymous login, personalized services. Personalized services include: (1) after users register, according to their registered information analysis, from the age, interest, occupation, as well as their access records in the system and other aspects of comprehensive consideration, provide personalized services to individuals. Recommend him credible web education resources for his reading. (2) Users who log in anonymously are also interested in other kinds of resources. This part will use the association rules in the database.

#### (4) Resource retrieval

The first mock exam module is divided into two parts: precise search and semantic search. Accurate search is for the input keywords, keyword matching, to achieve the search results. Semantic search aims at providing users with resources with high similarity after semantic similarity matching. Give users different choices. In advance, we classify the eight attributes of web education resources. When we search resources, we can reduce the scope of the search. We can select one of the attributes to search, or we can select several attributes or all at the same time. It also includes sub module quick download window module, which is a convenient way to download. It can be directly linked to the download function of resources, and there is no need to open the resources to download again.

#### (5) File upload

Mainly for users to reduce the phenomenon of free riding, so that resources can be shared more effectively. When a certain kind of users have a deep research or summary of which aspect of web education resources, this provides the upload module of trusted resources, so that the knowledge in a certain field can be trusted and efficiently present the same kind of users. At the same time, it also gives a general summary of different users, so that other users do not need to collect a large number of data for analysis and summary. This saves a lot of time and improves efficiency.

### 3.3 System module function

The main purpose of this module is to realize the personalized needs of users. Users can make personalized customized services for semantic based web education resources through members or anonymous login. For example, user a has a strong interest in economics. Just input keywords in the corresponding search box or in the process of information input, and the system can provide economics related content or hyperlinks. At the same time, user a will automatically push the latest news about economics to the user after the next login.

#### (1) Resource retrieval

This module is divided into two functions: one is to view resources, the other is to download resources. First, we search resources, which are divided into precise search and semantic search. Accurate retrieval is to match the keywords input by users accurately in the database, so as to get the corresponding web education resources.

Semantic retrieval is a formal calculation of semantic similarity for the keywords entered by users, infers the information content that may be related to the keywords, and gives them together, reflecting the intelligent search.

#### (2) File upload

Web education resources are constantly updated, so it is not enough to only rely on us to update the amount of data, so users need to inject new "blood" into us to realize the sharing mechanism of resources. After users get web education resources from the system, they can digest, absorb and summarize the trusted education resources, which can be uploaded to more users of web education resources to share. At the same time, we will also use the points management mechanism to give resource providers certain points reward through the access rate and download rate of file upload.

#### (3) Points management

In order to prevent the free riding phenomenon, we use the points management system to manage the users' behavior of uploading and downloading resources. In order to enable users to provide resources for each other to share, but also for the system to constantly update resources to contribute.

#### (4) User registration

In the first mock exam, we will analyze the information of user registration and give the resource push plate. For example, through the user's age, the user's industry, the user's interests and so on, the semantic analysis is carried out to get the information feedback to the customized section of the service, so as to realize the personalized service.

#### (5) Resource push

It mainly aims at the data statistics and analysis of all users' access records, and gets the relatively high reliable and high-quality Resource Recommendation in this system. Users can easily find the resources they need. Ranking is based on the number of visits to users, that is to say, through quantitative processing.

System database design and main interface

### 3.4 Main interface of prototype system:



Fig 4: Main interface diagram

## 4. Conclusion

In the increasingly large-scale Web education resources, how to find the highly credible Web education resources that are suitable for users' cultural background, personal interests and research fields has deep practical significance. web educational resources are shared, intelligent and cross-platform, which makes it more difficult for users to get the resources they need. To solve this problem, this paper first puts forward an agent-based pre-selection hierarchical model ERHMA for web educational resources, which screens the target resources twice: the first layer is classified and screened by the filtering algorithm based on semantic similarity; In the second layer, the filter algorithm based on Q learning is used for secondary screening to reduce the scale of candidate services. Then, the selected Web education resources are Ranked by rank again based on fuzzy sets and RSS. Firstly, the algorithm uses Euclid ambiguity as a measure of the degree to which the query content can vaguely represent web educational resources, and aggregates highly credible web educational resources in their own related fields through RSS technology. In this way, the web educational resources needed by the end users of web educational

resources can be ordered ahead of time, and the high-quality resources "needed by users" can be pushed for users.

### References

- [1] Zhu, & Jia. Midi and web teaching in colleges of multimedia system application. *Key Engineering Materials*, vol.1, no. 474-476, pp. 1926-1930, 2011.
- [2] Shinohara, Hideo. A study of teaching design in classroom web instruction : focusing on learning process and teaching system. *Journal of Hokkaido University of Education*, vol.1, no. 45, pp. 187-197, 1995.
- [3] Lyster, & Norman, C., The use of the web operating system to supplement the teaching of cobol. *ACM Sigcse Bulletin*, vol.18, no. 4, pp. 46-49, 1986.
- [4] Oliveira, & A. web teaching as culture: introducing the pontes approach. *International Journal of web Education*, vol.23, no. 3, pp. 205-216, 2005.
- [5] Ilhan Özgül. An analysis of the elementary school web teaching course in turkey. *International Journal of web Education*, vol.27, no. 2, pp. 116-127, 2009.
- [6] Boulet, & Marie-Michele. Designing and developing an intelligent advisor system for transfer tasks in web. *Computers & Education*, vol.19, no. 4, pp. 341-357, 1992.
- [7] Hodge, G. M., Jupp, J. J., & Taylor, A. J. . Work stress, distress and burnout in web and mathematics teachers. *British Journal of Educational Psychology*, vol.64, no. 1, pp. 65-76, 2011.
- [8] Yang, L., Ketner, K., Luker, S., & Patterson, M. . A complete system for publishing web-related etds. *Library Hi Tech*, vol.34, no. 1, pp. 151-163, 2016.
- [9] Hwong, N. C., Caswell, A., Johnson, D. W., & Johnson, R. T. . Effects of cooperative and individualistic learning on prospective elementary teachers' web achievement and attitudes. *Journal of Social Psychology*, vol.133, no. 1, pp. 53, 1993.
- [10] Maganioti, A.E., Chrissanthi, H.D. Cointegration of Event-Related Potential (ERP) Signals in Experiments with Different Electromagnetic Field (EMF) Conditions. *Health*, vol.12, no.2, pp.400-406, 2016.